

## Pattern, area and diversity: the importance of spatial scale in species assemblages

J. Podani<sup>1</sup>, T. Czárán<sup>1</sup> & S. Bartha<sup>2</sup>

*1 Department of Plant Taxonomy and Ecology, L. Eötvös University, Ludovika tér 2, H-1083 Budapest, Hungary*

*2 Institute of Ecology and Botany, Hungarian Academy of Sciences, H-2163 Vácraátót, Hungary*

**Keywords:** Information theory, Plot size, Sampling, Spatial analysis, Vegetation.

**Abstract:** Available methodology to detect pattern in plant communities (or sessile animal assemblages) is reviewed and criticized. Attention is focused on the area dependence of spatial analysis, as based on plot sampling. The discussion of procedures follows the conceptual scheme of Juhász-Nagy showing that the methods reflect different levels of understanding community characteristics. It is pointed out that Juhász-Nagy's information theory-based approach represents the most sophisticated analytical tool for the presence/absence case. We emphasize that any study of pattern-related vegetation phenomena should consider the spatial dependence of observation, measurement and description of vegetational properties.

### I. Introduction

Community-level pattern attracts considerable and steadily growing interest in terrestrial plant ecology. However, different aspects of pattern have received an unbalanced treatment in the literature, and the presentation of the topic is often misleading. Very often, the scale of pattern has been detected for separate species, with results published under the headline 'community pattern', forgetting that the latter is not merely a 'sum' or 'average' of single species patterns. Multivariate procedures offer at first glance a more adequate approach to this problem. Clustering and multidimensional scaling do reveal various structural properties of vegetation which relate to pattern (clusters, gradients, cf. Orlóci 1988), but scale (areal) dependence of classifications and ordinations is far too often neglected in most applications. It was Juhász-Nagy (1967, 1976, 1984, 1993; see also Juhász-Nagy & Podani 1983) who recognized first that analyzing the scale of community pattern requires a well-founded conceptual basis and a specific methodology utilizing appropriate

sampling design. Juhász-Nagy's approach and other methods developed for similar purpose, not smoothed into widely followed and fashionable trends in ecology, did not receive sufficient attention. The present paper tries to emphasize the importance of the topic through a review of concepts and available procedures, presents examples and outlines future perspectives.

### II. Characteristic areas

Plant communities are located in the 3-dimensional real world, yet community pattern is usually described as if the vertical dimension were non-existing or irrelevant. We do not engage in any dispute whether this simplification is correct or not, but realize that communities are most often treated as essentially two-dimensional entities. This is implicit whenever sampling design is characterized by specifying (two-dimensional) plot size, (planar) shape and arrangement (on the plane) ignoring vertical extension and any stratification. Further reduction of dimensionality to one, a practice in gradient analysis and in the

study of repetitive pattern, stationarity and heterogeneity along a fixed direction, is also common in ecology (see e.g., Dale, 1993, in this volume), but the present paper is concerned only with the two-dimensional aspect.

Given this startpoint, community variation is to be considered over *area*, and the spatial scale at which this variation is observed (or measured) is expressed in *area units*. Distinguished points on this scale, pertaining to some extreme values of measures of community variation, will be termed *characteristic areas* after Juhász-Nagy (1967).

### II.1. Simple examples

The area dependence of community variation is demonstrated with an artificial 3-species community in Figure 1a. The example is deliberately very simple: plant pattern is completely random, individuals are considered point-like in a continuous space and the species have approximately equal densities. At small area (Fig. 1b) plots are likely to include only a single species or none; the data obtained do not provide any useful informa-

tion (in the common sense of the word) regarding the joint behavior of species in the community and the variation covered is no more than a list of species present. At the other extreme of the scale, every plot of very large size will include all the species (Fig. 1d). A presence/absence data matrix for this case exhibits no variation and is again uninformative on multispecies pattern. The intermediate size shown in Fig. 1c will yield a more diverse sample: some plots still have a single species but many include various combinations of the three species. This is approximately the scale point at which plots seem to be maximally different, thus providing a lot more information on community variation than the other areas.

This "best" plot size is influenced by many textural and structural properties (*sensu* Barkman 1979) of the community (abundances, aggregation of individuals, positive or negative species associations, etc.). Let us examine just one more "community" with a pattern markedly different from the previous example (with aggregated spatial dispersion of species) but abundances remaining the

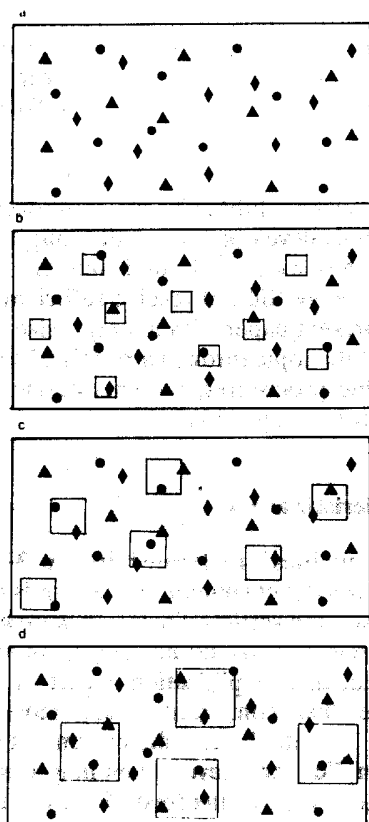


Figure 1. The effect of plot size on the species composition of plots in a 3-species random assemblage.

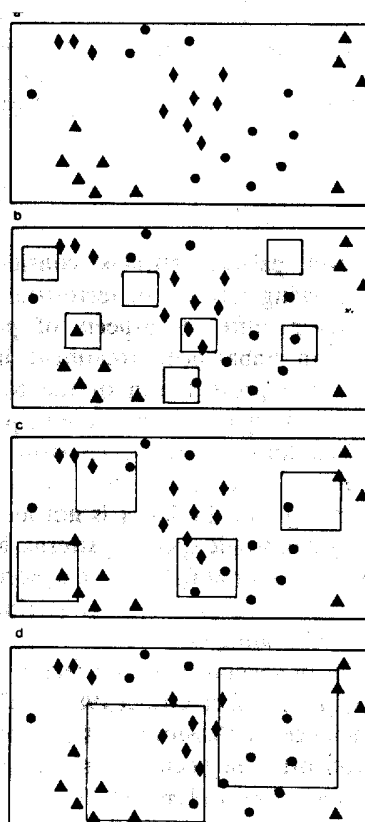


Figure 2. The effect of plot size on the species composition of plots in a 3-species community with a definitely non-random plant pattern.

same (Fig. 2a). The area which proved the most informative in the previous case is now too small: only a single species occurs in most plots (Fig. 2b). We need larger plots to capture a scale point leading to a much more "diverse" sample (Fig. 2c). The plot size at which all plots include all the species is apparently larger than in the previous example (Fig. 2d).

## II.2. An earlier concept

An interesting historical fact deserves attention at this point. Of the various approaches to community ecology, the Gestalt-oriented Braun-Blanquet tradition, which had very little affection to numbers and statistics, showed the greatest interest in areas that can be used in community characterization. The so-called *minimum area concept* of phytosociology refers to a plot size at which "typical characteristics of the community start to develop" (see Whittaker 1973, Barkman 1989, for review). The need for finding this area was desperate, but definitions were at best inoperative (or, at worst, circular). Most often, species/area curves were used to find a plot size beyond which increases in the total number of species become negligible. Despite the considerable effort devoted to the search for "flattening" or "inflexion" points in species/area curves, the approach failed.

## II.3. Estimation vs pattern detection

Even though different species/area models assume certain underlying distributional properties of the community, scale of pattern remains unexplained by these curves. As Bouxin & Gautier (1982) note: "one may wonder how the traditional species-area curve can serve to define a minimal area when the curve is a result of a mixture of very different patterns". The explanation lies in the unavoidable distinction between two contrasting objectives of vegetation sampling: *estimation* of some population (or community) parameter and *pattern detection* (Kenkel et al. 1989). As they point out, for estimation objectives reduction of sampling variance is crucial, whereas pattern detection requires maximized sampling variance (max. heterogeneity of data). The species/area relation has to do with estimation: upon area increases we achieve an increasing precision and accuracy of species number for the whole community. Similar is the dependence of "species-diversity" upon area (Pielou 1975, p. 12) and many other estimates that are used in conventional statistics (mean cover of

a species, etc.). A common feature in such investigations is that the larger the area sampled (via either plot size or empirical sample size increases) the better the estimation, and convergence to the theoretical value is expected (supposing that estimates are not biased).

The relationship of structural, pattern-related properties to sampling design is different: plot size and empirical sample size increases do not produce the same effect. Measures that can depict area dependence of community variation (i.e., the scale of multispecies pattern) respond to plot area increases in another way: instead of converging to a theoretical value, they exhibit some peaked effect where variation is the highest. In a full analogy to the analysis of univariate pattern, characteristic areas of communities can be defined in terms of maxima and minima of meaningful measures of multispecies pattern. Such areas correspond with the scale of multispecies pattern, just like plot (or block) sizes associated with extreme values of the variance/mean ratio for a single species. It is therefore inevitable that a series of plot sizes is required in sampling, so detection of scale of community pattern is a case of *space series analysis* (Podani 1992). Sample size increases, on the other hand, have the same effect as for population parameters: the more plots (of a given size) are examined the better the estimates of measures of community variation and characteristic areas. Thus, estimation is also involved in this case, but the dependence of estimates on area is of primary concern.

## III. Admissible sampling designs

Before discussing measures that meet the above requirement, we discuss basic types of sampling designs that can be used for evaluating scale dependence of community variation. Of the four basic characteristics of plot sampling (i.e., number, size, shape and arrangement) plot size is to be increased, whereas the others should be kept constant, to avoid confounding effects of changing two or more sampling characteristics.

### III.1. Plot shape

It is preferably isodiametric (circular or quadratic) so as to minimize vegetational heterogeneities (edges) captured by a single plot. Long sampling units are obviously inadequate for evaluating areal aspects of scale, because they are more likely to

run through more diverse parts of the community and to include plants within the same plot even if they actually grow far apart in the field<sup>1</sup>.

### III.2. Arrangement of plots

Some random component in the arrangement of plots is generally recommended to reduce the chance of biased results. Plots located at preferred sites may also be used, although results obtained from such a sample cannot be generalized for the whole community. Full randomization is difficult to achieve in the field (although it can be done easily in computer simulated sampling experiments). Randomization and computation of measures of community variation may be done separately for each plot size. Its advantage is that samples can be considered independent. A nested arrangement of increasing plot sizes based on a single arrangement is also possible (Fig. 3), to facilitate less laborious field sampling. Systematic arrangement of very small contiguous square units in a grid randomly located in the community is also possible, so that sampling effort can be greatly reduced. Data for larger plots can be generated by successive fusions of neighbouring units. Care must be taken to keep plot shape constant, so that the confounding effects of plot size increases with alternating shapes of rectangular and quadratic blocks (as in the Greig-Smith technique) are avoided (Pielou 1977). However, in this case the analysis is restricted to a small portion of the community and generalizations to the full community should be done with caution (replicate grids may provide a solution).

In an ideal situation sample plots do not overlap one another, so that the data set thus obtained can be subjected to significance tests, if other sampling criteria do not violate the independence criterion. In the nested arrangement the samples will depend on each other and in computerized sampling (Podani 1987) the studied area is usually "oversampled" such that small parts of the study region may get into many sample plots. In these cases conventional significance tests do not apply. However, for pattern detection objectives independence is not always a strict requirement. For descriptive purposes, e.g., in detecting characteristic areas, oversampling may provide better es-

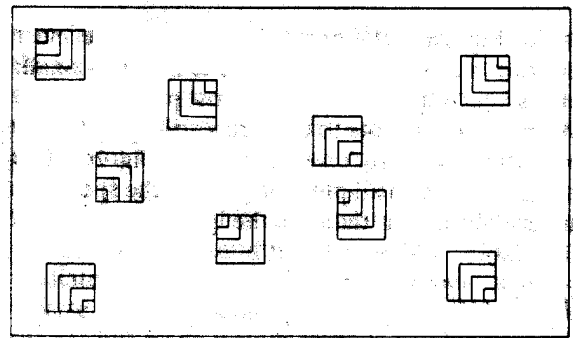


Figure 3. Random arrangement of nested plots for studying multispecies pattern in a community.

timates than the traditional strategy because of the much larger empirical sample size that can be used. If one insists to attach some probabilities to the resulting figures to allow comparisons, the bootstrap technique (help yourself via resampling your own data thousand times!) is recommended<sup>2</sup>.

### III.3. Empirical sample size

We shall use the term sample size in the statistical sense: technically sample size refers to the number of sample plots or sampling units, and sample will refer to the full set of such plots. (In ecology, a sample is often understood as a single sampling unit, and this is a common source of confusion.) As mentioned earlier, sample size increases affect the estimation of measures of community variation and characteristic areas. The analytical techniques greatly differ in sample size requirements. A relatively large sample is necessary for detecting characteristic areas associated with the collective behavior of species (see subsection IV.3), whereas conventional samples of size 50-200 are usually sufficient for the other strategies (subsections IV.1-2).

### III.4. Plot sizes

There is no general rule as to the selection of the smallest and the largest size in the series: the range should be broad enough to capture the characteristic areas but these areas are known only *a posteriori*. Nevertheless, plant and community size and our previous field experience should govern our choices. The changes at each step of the series are also arbitrarily defined. These chan-

- 1 Long units obtained by fusing transect cells are used in the analysis of direction-related phenomena of vegetation pattern.
- 2 Edge effects may also be critical if sample plots are relatively large compared to the area of the study region. Since overlaps with the boundary line are not allowed, in random arrangements the central part of the region will be more intensively sampled than the peripheral parts. The usual toroidal corrections for edge effect do not work, however, because fusions of remote fractional plots are meaningless for the analysis of multispecies patterns (cf. Podani 1987).

ges should preferably be as small as possible to be able to detect the characteristic areas precisely, but time and cost limitations will dictate our choice.

### III.5. Data

Species presence, cover or number of individuals of species is recorded for each plot at each size, to yield a series of data matrices,  $X_1, X_2, \dots, X_k, \dots, X_p$ , where  $p$  is the number of different plot sizes. The size of each matrix is  $n$  by  $m$ ,  $n$  is the number of species detected in the community,  $m$  is the number of plots (sample size) kept the same for each plot size. In  $X_k$ , summation over rows for column  $j$  is denoted by  $x_{.jk}$  (plot total), summation over columns for row  $i$  is  $x_{i.k}$  (species total).

### IV. Approaches to detect scale of multispecies pattern

Part of the conceptual basis developed by Juhász-Nagy is the sequence of certain objects and operations associated with vegetation surveys (see e.g., Juhász-Nagy 1993, Figure 4). The basic scheme of flora  $\rightarrow$  vegetation (or in a general context: basic set  $\rightarrow$  compositional structures) is refined as point set (flora, fauna, biota)  $\rightarrow$  simplex set (e.g., abundances of elements)  $\rightarrow$  Venn complex (showing interaction of components, e.g., species associations)  $\rightarrow$  sorted complex (such as an ordination)  $\rightarrow$  allocated complex (e.g., a vegetation map). Each step represents a different level of understanding community characteristics and different details. Detection of the community-level

pattern can be attempted at the first four stages along this sequence. The subsequent discussion of available methodology will follow the above scheme.

### IV.1. Methods related to the basic set

For vegetation surveys the basic set is composed of plant species that are present in the community studied. The methods will operate on species subsets found in sampling units; such a subset in a sample plot will be termed the *florula* (Juhász-Nagy & Podani 1983) to make distinction from *flora*, a universal set in this context, which is the collection of all species present in the community.

#### IV.1.1. Number of species and its variance.

In the simplest case the size of florulas, i.e., the number of species ("species richness",  $x_{.jk}$  for presence absence data) per plot, is considered as a starting point for analyzing community variation. Relevant published work stems from Pielou's (1972) pioneering studies. She suggested to use the *difference* between the variance of species number,  $\text{Var}[x_{.jk}]$ , and the expected variance under the null hypothesis of completely independent distribution of species given by

$$E(\text{Var}[x_{.jk}]) = \sum_{j=1}^m \frac{x_{.jk}}{m} - \frac{x_{.jk}^2}{m^2}$$

as a measure of overall association of species. Robson (in the Appendix to Pielou 1972), derived a test statistic in form of the ratio of the actual and expected values, which was subsequently used by

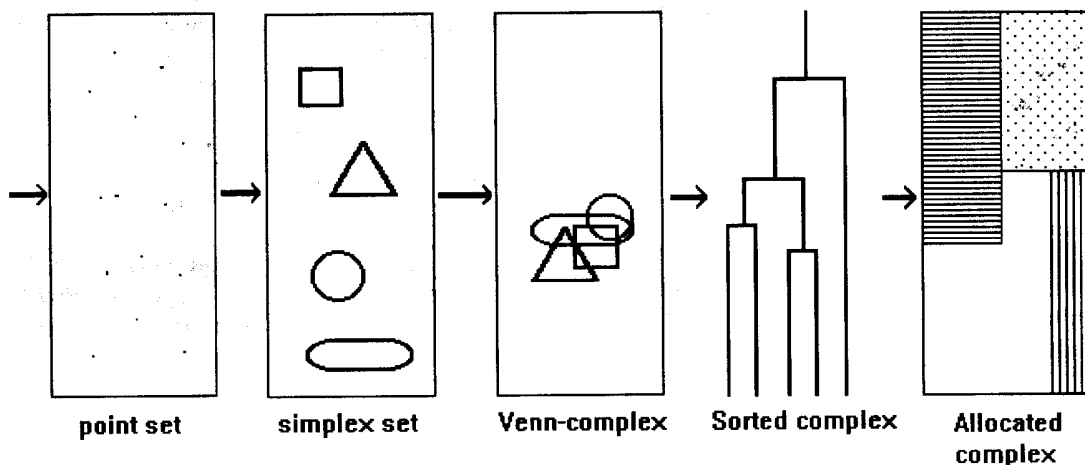


Figure 4. A conceptual scheme illustrating different levels of understanding vegetation characteristics (after Juhász-Nagy 1993).

several authors (e.g., Schluter 1984, McCulloch 1984, Palmer 1987, Wilson et al. 1987, Watkins & Wilson 1992). It is interesting historically that Pielou completely neglected the spatial aspect of this statistic, and demonstrated its use on a single plot size in a succession study, notwithstanding that spatial dependence of species associations had long been known (cf. Greig-Smith 1985). Palmer (1987), for example, examined the dependence of  $\text{Var}[x_{jk}]$  and  $\text{Var}[x_{jk}] / E(\text{Var}[x_{jk}])$  upon plot size in oldfields and detected maxima in several cases, whereas monotone increases in others. His results were inconclusive as to the overall behavior of this function probably because of the relatively few scale points (4) and narrow scale (0.01 to 10 m<sup>2</sup>) that he used. Wilson et al. (1987), using an even narrower range of plot sizes (up to 1 m<sup>2</sup>), found "little evidence" of plot size dependence. This led Zobel et al. (1993) to rely on a single plot size in a succession study followed by a detailed and very sophisticated statistical analysis of data. We feel, however, that complete ignorance of scale dependence in such surveys is quite dangerous and potentially misleading (see also Kenkel et al. 1989). At the same time, the same authors (Zobel et al. 1993, p. 490) correctly point out that interpretation of changes in this ratio by background ecological phenomena (e.g., niche limitation, positive segregation, etc.) is not unique (see also Lepš 1990a, for similar comments on single species patterns and interspecific associations). We can agree completely with Schluter (1984, p. 1003) who goes even further by saying that "there is no necessary correspondence between the result of the variance test and any ecological process. Researchers frequently seem want to infer the existence of a particular interaction between species (e.g., competition) on the basis of a statistical test... but such inference is usually not valid". Computer simulations, with known population and community parameters, may be helpful in this regard. For example, Tóthmérész & Erdei (1992) showed the dependence of  $\text{Var}[x_{jk}]$  upon plot size as well as the dominance relations between species and found definite peaked effects. Obviously, more field data and simulations are required to examine this, actually the simplest possible measure of community variation.

#### IV.2. Methods using simplex sets

A typical simplex set comprises frequency distributions of species; often used in phytosociology (coenological spectra) and termed the *texture* by

Barkman (1979). Functions to be discussed in this subsection will be based on such distributions.

##### IV.2.1. Variance of species-diversity.

Pielou (1966) suggested the term "pattern diversity" which is "high when the individuals of the various species are thoroughly mingled so that several species are usually present in any small sub-area; it is low if the species are segregated so that small sub-areas are likely to contain individuals of only a few of the species." This concept was first used to n-tuplets of species (e.g., nearest neighbors), rather than sample plots, but later it was applied to plot data as well (e.g., Zobel et al. 1993). The computations involve the determination of species-diversity in each plot (using the Shannon-Weaver entropy-estimator with natural logarithm), denoted by  $H'_{jk}$ , which is potentially replaced by  $\exp(H'_{jk})$  to measure the number of equally common species that would produce the same  $H'_{jk}$  as the actual data (Peet 1974).  $\text{Var}(H'_{jk})$  is then used to measure pattern diversity. The expectation of this variance for the null situation is unknown, but Zobel et al. (1993) used the bootstrap technique to obtain significance levels for the deviations of  $\text{Var}(H'_{jk})$  from the expectation. Whether or not  $\text{Var}(H'_{jk})$  and the deviations exhibit peaked effect over area increases requires future investigations.

##### IV.2.2. Local distinctiveness (pooled entropy).

The concept of entropy allows us to define an alternative spatial characteristic of communities: the plot size with the highest uncertainty that any randomly selected plant species is found in the plot. For small plots this uncertainty is relatively low; we can be pretty sure that a randomly chosen species will be absent. On the other extreme, for large plots, there is also a low uncertainty because most species are likely to be included. There is an intermediate plot size for which this uncertainty is the maximum, where the separate species are the most distinctive among the plots. Two plots, on the average, differ in the greatest number of species at this size. Hence the name *local distinctiveness* suggested by Juhász-Nagy, but other terms also appear in the literature of ecology and taxonomy (e.g., "total information" or "information content"). Mathematically it is perhaps most straightforward to use the term *pooled entropy*; because this measure is additive, being obtained as the sum of single species entropies:

$$m\hat{H}_{ik}([L]) = \sum_{i=1}^n m\hat{H}_{ik}$$

where

$$m\hat{H}_{ik} = m \log m - x_{i,k} \log x_{i,k} - (m-x_{i,k}) \log (m-x_{i,k})$$

is the entropy of species *i* at plot size *k* (recall that it is present in  $x_{i,k}$  of *m* sampling units). Graphically, local distinctiveness is conceived as the sum of areas of plane figures in a Venn-diagram. Each species has its own maximum area of entropy; its dependence on abundance and spatial dispersion is demonstrated in Figures 5-6, respectively, in simulated "populations". The maximum point for  $m\hat{H}_{ik}([L])$  is called the *compensatory area*<sup>3</sup> (Juhász-Nagy & Podani 1983). Its dependence on plot size over time, from a case study on primary succession, illustrates the scale problem commonly disregarded in succession studies. For small plot sizes (up to 20 m<sup>2</sup>) local distinctiveness and the compensatory area decreases over time (Fig. 7), whereas for large plot sizes (cca 30 m<sup>2</sup> and more) the temporal trend is the opposite. Working with small plots we detect increasing heterogeneity (divergence) whereas with very large plots we would find decreasing heterogeneity (convergence) during the same period. An investigator

who happens to select a plot size at about 20 m<sup>2</sup> would not detect any change over succession, and we do not have to emphasize how misleading con-

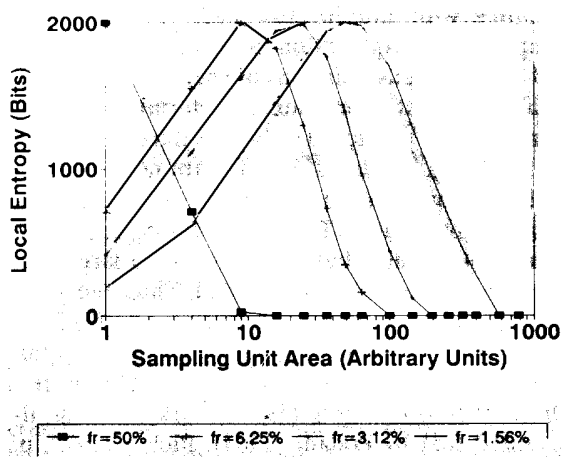


Figure 5. Plot size dependence of species entropy in four random populations with frequencies ranging from 1.56 to 50% (frequency is proportional to commonness).

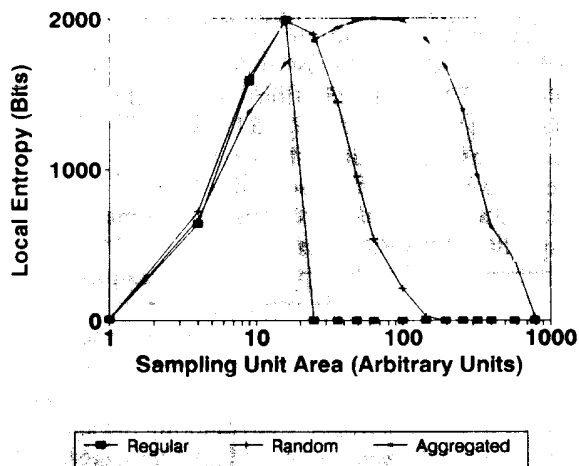


Figure 6. Plot size dependence of species entropy in three simulated populations with different spatial dispersion (with a constant frequency of 6.25).

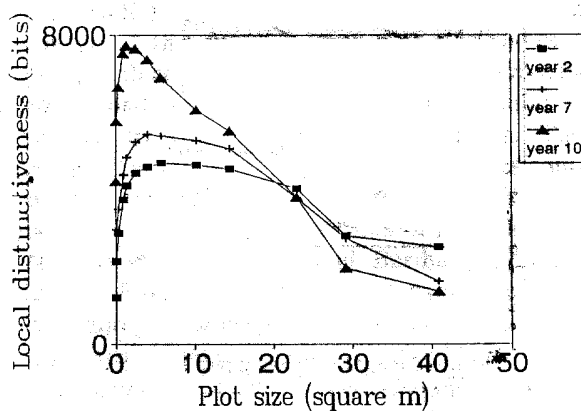


Figure 7. Plot size dependence of local distinctiveness over succession (from Bartha 1991).

- Williams et al. (1969) suggested a method called *multiple nearest neighbor analysis* which does not use sample plots but deserves attention in this paper. Instead of plots, the position of every plant individual is recorded and all inter-plant distances calculated. Each plant, together with its *q* nearest neighbors is considered as a "sampling unit" so that we have as many units as plants. In this case *q* is changed in the series. Williams and his co-workers computed the pooled entropy for this clump system and found that it is entirely the function of the mean distance from the reference individuals, i.e., it is distance and consequently area that matters rather than *q*. Thus, a *hybrid* procedure of pattern analysis may also be conceivable. In this, each individual may be considered as the centroid of a circular plot with radius *r*, and all species found within this plot are recorded. In this case the change of *r* would represent a series of plot size increments. The radius at which local distinctiveness reaches maximum will be analogous to the compensatory area.

clusions could be drawn from such an observation (see Bartha 1991).

#### IV.2.3. Entropy of marginal distributions.

Based on presence/absence data, Juhász-Nagy (1976) has suggested to apply entropy measures to the marginal distributions in the data table for each plot size, also discussed by Camiz (1993, this volume). These distributions are local valence, containing the  $x_{i,k}$  values (and local invulence containing  $m - x_{i,k}$ ) which show how many times a species is present and absent, whereas the floristic valence and invulence (containing  $x_{j,k}$  and  $m - x_{j,k}$ , respectively) reflect species richness of plots. The Shannon-Weaver entropy estimates applied to valences monotonically increase over plot size, while the entropy estimates of invulences decrease (Fig. 8). The small range of plot sizes, defined by the intersection of entropy curves may be considered as a *characteristic interval* along the spatial scale for that community. Many characteristic

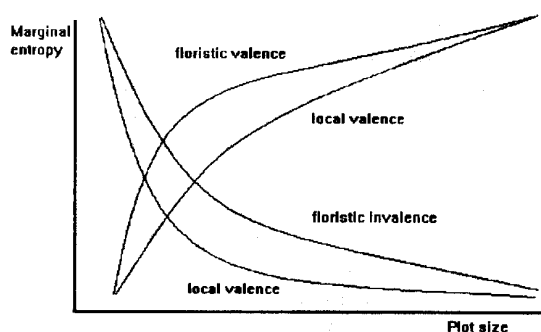


Figure 8. Idealized diagram showing plot size dependence of marginal entropy estimates.

areas (mentioned above and to be discussed later) are expected to lie within or at least close to this plot size range (Juhász-Nagy 1984).

#### IV.3. Methods based on complex sets

The previous sections considered the behavior of each species independently. One step ahead along the conceptual sequence involves measurement of *interaction* between components of the basic set (e.g., association between species or resemblance between sample plots). Associations may be restricted to a pair of species (most commonly) or,

as we shall see in the discussion of Juhász-Nagy's most original models, to any subset or even the complete flora. Central to Juhász-Nagy's models is a new concept of diversity applicable to supraindividual objects, such as florulas, which has a clear relationship to association.

#### IV.3.1. Number of species combinations (florulas).

From the series of plots used in the artificial examples it is seen that at small size very few combinations of species (*florulas*) are manifested (no species or either of the three). Increasing plots will capture more different combinations, but beyond a point the plots become "saturated" with species. It means that the plots will tend to include all or nearly all species of the community, so that the number of species combinations will necessarily decrease. If the plots are large enough in the last sample of the series, all of them will include every species. The existence of a maximum in this series is a mathematical fact without direct ecological reasoning.

Of course, all the possible combinations (i.e.,  $2^n$ , including the "empty" florula) will not appear in the sample, even if sample size would allow that (except the theoretical situation of random communities with infinite size and infinitely large sample). The explanation is textural constraints and the spatial associations of species. Mostly textural constraints dominate in determining actual combinations (up to 80-99%, see Bartha 1992 for explanation). If species A and B are negatively associated, combination {A,B} is very unlikely to occur in the sample. They may be included in the same plot, though, but this must be so large that other species will also be included. Thus, the *number of species combinations* or florulas (NSC) is useful at first glance to characterize some overall compositional behaviour of species. This is implicit in Juhász-Nagy's (1967) work and was illustrated first by Podani (1984). Pielou (1972, p. 338) also mentioned these combinations but without further use in her work. The area at which the maximum is obtained may be termed the *maximum area of species combinations*. Podani (1984) examined this function in detail by comparing field situations with randomly simulated counterparts. Bartha (1992) described the changes of the maximum for NSC over succession and found that difference between random expectation and actual values increases with time: inappropriate (or ecologically impossible) combinations do not appear in the sample. As he showed, increases in species number do not necessarily lead to an in-



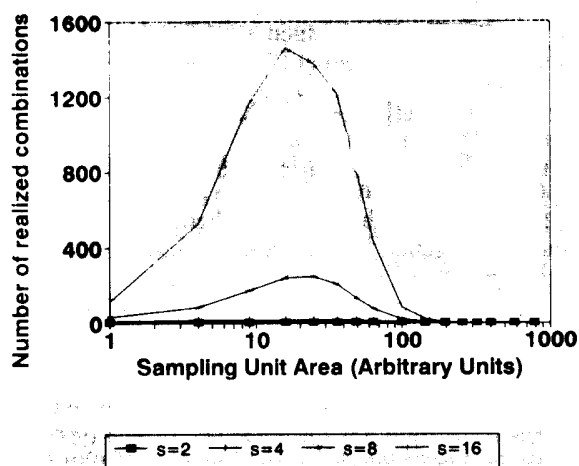


Figure 9. Plot size dependence of the number of realized species combinations in four random assemblages with different numbers of species (2-16).

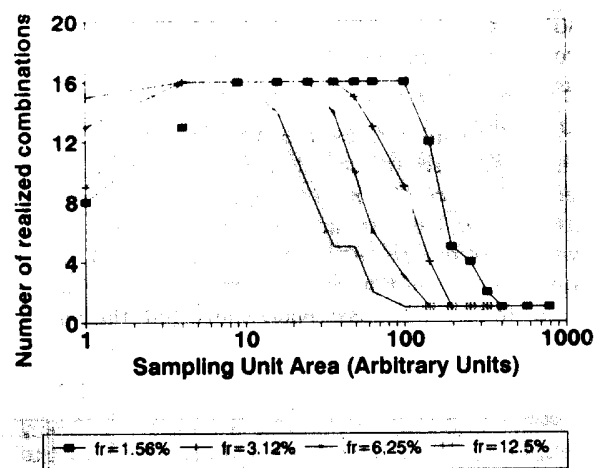


Figure 10. Plot size dependence of the number of realized species combinations in four different random assemblages with species frequency ranging from 1.56 to 12.5%.

crease of NSC as succession proceeds. In a simulated example (Fig. 9) with 2, 4, 8 and 16 species randomly dispersed we demonstrate the opposite effect: NSC increases as species number increases, although the maximum area of species combinations remains fairly the same. Thus, the null situation and the realized NSC-s may differ considerably, the latter is lower in reality. In another example (Fig. 10) we show the dependence of NSC on the commonness and rarity of species (i.e., texture), based on random spatial distribution for all species. Frequency (commonness) of species and the size of maximum area are negatively related, the higher the frequency the smaller the area of maximum NSC.

#### IV.3.2. Florula diversity.

The NSC measure does not distinguish between rare and common species combinations; they are equally considered. In other words, NSC may be the same for a sample in which all realized combinations equal in number and in another sample dominated by a single combination. Such a difference becomes detectable if we apply a diversity measure to the frequency distribution of florulas. With a full analogy to species-diversity, Juhász-Nagy (1967, 1976, 1984, see also Juhász-Nagy & Podani 1983) suggested to use Shannon's measure to express florula diversity of the community. (In this case the basic unit is the florula of a sample plot rather than a plant individual, and the categories are species combinations rather than species.) The formula of florula diversity is given by

$$m\hat{H}_k(A,B,\dots,N) = m \log m - \sum_{s=1}^{2^n} f_{sk} \log f_{sk}$$

where  $f_{sk}$  is the frequency of the  $s$ th species combination in the sample,  $m$  is sample size,  $k$  refers to plot size, and  $A,B,\dots,N$  refer to  $n$  species. Mathematically,  $m\hat{H}_k(A,B,\dots,N)$  is the joint entropy of species in the  $X_k$  data matrix. Graphically, this is the envelope of the Venn complex (Fig. 4). The change of this function over plot size shows a peaked effect: *the maximum area of florula diversity corresponds to a plot size where the community reaches its richest compositional pattern.* The chance that two plots have the same species composition is minimized at this scale point. It may be fairly close to the maximum area of NSC. Juhász-Nagy & Podani (1983) and Bartha (1992) examined florula diversity changes over succession concluding the same as in case of local distinctiveness. Tóthmérész & Erdei (1992) observed in a simulated sampling experiment that florula diversity and the variance of the number of species per plot tend to indicate the same scale in certain types of pattern. Curves of the variance were less smooth than florula diversity curves, showing that florula diversity is not in a one-to-one correspondence with inter-plot differences in species number. They also showed that potential bi- or trimodality of the diversity curve, i.e. several levels of scale, is caused by the existence of distinct dominance groups (e.g., very rare and very common species) in the community. Here we illustrate the effect of the number of species (Fig. 11) and abundance (Fig. 12) on the change of florula diversity in simulated random communities.

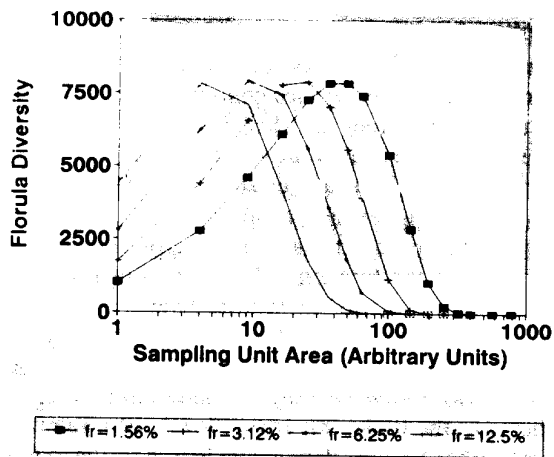


Figure 11. Plot size dependence of florula diversity in four different random assemblages, each with 4 species. Average species frequency ranges from 1.56 to 12.5 in the assemblages.

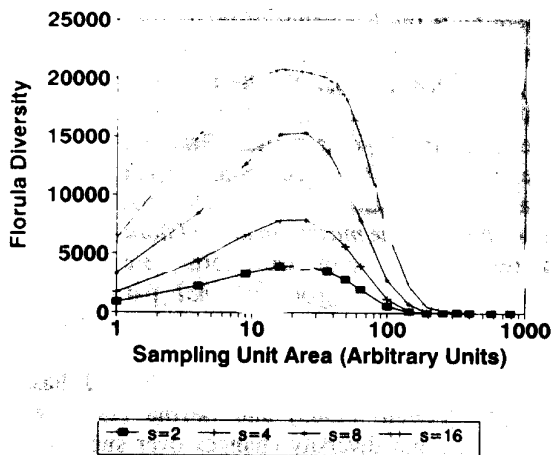


Figure 12. Plot size dependence of florula diversity in four different random assemblages, with species number ranging from 2 to 16.

IV.3.3. *Associatum*.

The dependence of interspecific association upon plot size has long been known in plant ecology (cf. Greig-Smith 1985, and many references therein). Two species may be negatively correlated at one scale point, positively at another, and may exhibit no correlation at some third point. For measuring the overall association of every species in the community, however, pairwise correlations and associations are less straightforward. An appropriate measure is the contingency information of the 2<sup>nd</sup> contingency table (Juhász-Nagy 1967), a quantity termed by Juhász-Nagy the *as-*

*sociatum*. This is obtained as the difference between pooled and joint entropy

$$m\hat{H}_k(L) = m\hat{H}_k([L]) - m\hat{H}_k(A,B,\dots,N).$$

A graphical illustration is the intersection part of the Venn-diagram. If species positions are independent of each other, the theoretical value of the contingency information is zero regardless plot size. This case of complete spatial independence can be approached as the sampled area and sample size tend to infinity in a random community. In any real situation, however, with limited areas sampled and relatively small sample sizes, *associatum* will show a peaked effect (*area of maximum associatum*) in the function of plot size even if the community is completely random, because of incidental positive and negative associations among rare species. This is obviously an artefact from an ecological point of view, therefore the deviation of an actual value from the *associatum* calculated for randomly simulated counterpart communities (null models) should be used.

Figure 13 illustrates the relationship between three measures of scale dependence in plant communities: *local distinctiveness*, *florula diversity* and *associatum*, based on an example from a secondary weed assemblage. As seen, the maxima of these functions do not necessarily appear at the same area. In fact, as Juhász-Nagy suggested, the ordering of maxima is also characteristic of the community (in this case the area of max. florula

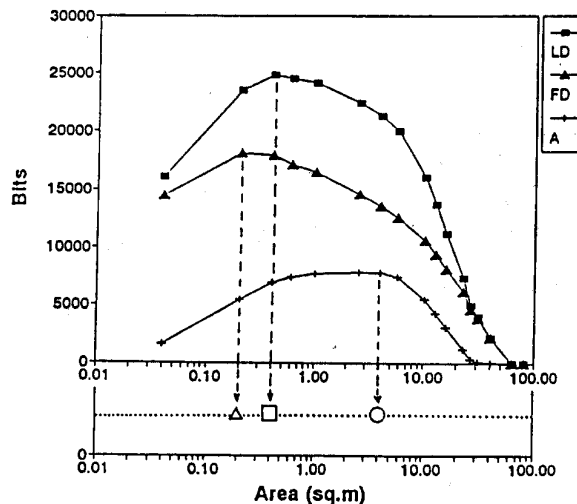


Figure 13. Plot size dependence of local distinctiveness (LD), florula diversity (FD) and associatum (A) in a 3-year old weed assemblage on a dump from strip coal-mining (Bartha 1993). Note the resulting characteristic ordering:  $A_{flor} < A_{comp} < A_{ass}$ .

diversity is the smallest, close to the compensatory area, whereas the area of maximum associatum is much larger).

Random references for comparisons with actual associatum values can be produced for any point pattern either by Monte Carlo simulation or by direct calculation of florula frequencies based on the assumption of Poisson-distributed and spatially independent species patterns. Both methods have their advantages and disadvantages. Direct calculation is fast and easy, since the spatial independence assumption ensures that  $m\hat{H}_k(L)=0$  for all  $k$ , which in turn yields  $m\hat{H}_k(A,B,\dots,N)=m\hat{H}_k(L)$  (Juhász-Nagy 1984). It is simple to compute local distinctiveness once the probabilities of presence in a sample plot of a given area are known for all species. But as mentioned above, the theoretical case, being the result of idealized postulates, will not usually realize in actual situations because of textural constraints, even if no spatial dependence exists.

On the other hand, Monte Carlo simulation of point patterns can be more realistic regarding the size and the shape of the area sampled, but it requires a lot more computing time, and the incidental effects of rare species' associations might produce a wide variety of outcomes, which is quite uncomfortable to use as a reference.

#### IV.3.4. Expected resemblance (mean floristic similarity).

The common problem with florula diversity and associatum is their large sample size requirement. Small samples are usually inadequate to faithfully represent frequencies of species combinations occurring at the given scale point, and the resulting florula diversity estimate may be invariant over  $k$ . If we do not wish to get a detailed analysis using information theory models, there is a simpler alternative way to express scale dependence of communities, based on the concept of expected inter-plot distance or *resemblance* (Podani 1984). The starting point is that maximum florula diversity implies maximum or nearly maximum inter-plot distances or dissimilarities. Therefore, an area for which inter-plot distances are maximum is a good approximation to the maximum area of florula diversity.

If we randomly locate random pairs of plots of a given size within the community, we can calculate a distance measure for each pair. A very large number of such plot pairs, each pair located independently from the others, will then provide us

an estimate for the mean inter-plot distance at the given plot size. Plotting the estimates over plot size summarizes the spatial dependence relations in the community (Fig. 14). A problem with such independent pairs is that they are difficult to specify in the field, although computer-simulated sampling may help if digitized point maps are available. For field data, a resemblance matrix between plots may be calculated for each plot size, and the average resemblance is calculated based on the off-diagonal elements (e.g., Wildi & Krüsi 1992). (This method is less elegant, because every plot contributes  $m-1$  times to the average.) The use of averages goes far back in the history of phytosociology; there were quite few attempts to define the minimum area of communities in terms of average similarity or within-community homogeneity (Gounot & Calléja 1962, Moravec 1973, Dietworst et al. 1982).

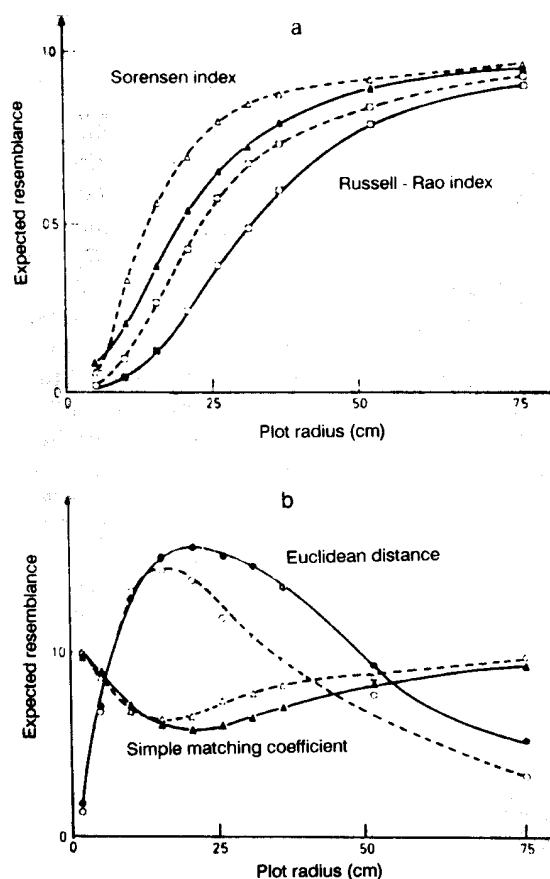


Figure 14. Expected resemblance in the function of plot size in a grassland community (solid lines) and its random counterpart (dotted lines, from Podani 1984). a: Indices disregarding value of  $d$  in the  $2 \times 2$  contingency table. b: Coefficients considering  $d$ .

A crucial decision in this approach is the choice of resemblance coefficient. The Sørensen index and related measures, which do not utilize cell *d* from the 2x2 contingency table, increase monotonically over area (Fig. 14a). Therefore, no objective method can be defined to reveal any characteristic areas from such curves which asymptotically approximate the maximum. In this direction we cannot go farther than with species/area curves. On the other hand, distance-related resemblance measures (e.g., Euclidean distance for binary data or the simple matching coefficient) will have some extremum, showing the area where the sample plots maximally differ (Fig. 14b). Simulation studies indicated that this distance approach identifies a characteristic area closely approximating maximum area of florula diversity (compare Figures 4 and 5 in Podani 1984). This approximation may be extremely useful, for example, in plot size optimization before clustering. For cover data, however, no such extreme points were found, the curves monotonically increased even for distances<sup>4</sup>. There is very little experience with the use of expected resemblance, however, and further studies are required to judge the usefulness of this approach<sup>5</sup>.

#### IV.3.5. Partitioning of eigenvalues.

This method was originally proposed by Noy-Meir & Anderson (1971). Covariance matrices ( $C_k$ ) are calculated for the species at each plot size *k* in a grid, the large plots obtained from fusions of small units, and then they are summed to get a total covariance matrix. After an eigenanalysis of this matrix, the eigenvalues,  $\lambda_i$ , are partitioned into the amount that each plot size contributed to that value based on the following equation:

$$\lambda_i = \mathbf{v}_i' C_k \mathbf{v}_i ,$$

where  $\mathbf{v}_i$  is the eigenvector pertaining to  $\lambda_i$ , normalized to unit length.  $\mathbf{v}_i' C_k \mathbf{v}_i$  is then plotted against block size. Peaks indicate block sizes contributing much to that eigenvalue, thus reflecting scale of multispecies pattern to the extent the eigenvalue accounts for the total variation. There is a separate curve for each eigenvalue, so that as

the authors suggest several independent scales of pattern may be identified from the data. A modification of this technique is due to van Hoef & Glenn-Levin (1989).

#### IV.4. Definitions based on sorted complexes

Sorted complexes are the conventional results of multivariate analysis, such as ordinations and classifications. Although not typical, there are some specific applications of multivariate procedures that allow detection of scale in community patterns.

##### IV.4.1. Analysis of three-way tables.

Camiz & Gergely (1990, see also Camiz, 1993, this volume) suggested to analyze all the *p* data matrices simultaneously by Escoufier's (1973) three-way scaling technique, STATIS. They found that the position of data tables in the ordination plane corresponds well with characteristic areas detected by the Juhász-Nagy models: maximum area of florula diversity was an extreme on a second axis, whereas maximum distances from the origin identify a plot size with maximum associatum. The proposition that extreme positions in these ordinations directly correspond to characteristic areas in every case needs confirmation, however.

##### IV.4.2. Pattern analysis of ordination scores.

The essence of this strategy is to reduce the "problem" with multivariate data to a univariate situation via ordination of sampling units in a transect. Species scores are thus replaced by the corresponding ordination score [Galiano (1983) used reciprocal averaging, Sterling et al. (1984) applied principal components analysis, Gibson & Greig-Smith 1986 used detrended correspondence analysis]. Then, conventional methods of univariate pattern detection (e.g., two-term local variance methods, Hill 1973) are used to quantify scales of community pattern. Galiano (1983) performed this analysis separately for the positive and the negative section of the ordination axes, giving two pattern analyses for each axis. Galiano's approach was strongly and rightly criticized by Gib-

4 Lepš (1990b) has suggested to use mean Euclidean distances in blocked-quadrat analyses of transect data for detecting community pattern.

5 Mean similarity is somewhat related to the concept of  $\beta$  diversity commonly used in ordination studies as a measure of species turnover along gradients: mean similarity is inversely related to species turnover (see Wilson & Shmida 1984, for review). Studies employing various plot sizes provide inconsistent results on the relationship between  $\beta$  diversity and area (cf. Økland et al. 1990), probably because of confounding effects of small scale pattern (detectable by changing plot size) and a larger scale pattern (apparent only along the gradient). However, the range of plot sizes used for such purposes does not allow direct comparison between the information theory functions and  $\beta$  diversity. Such an evaluation is surely an important task in the future in order to fill the gap between existing approaches to analyzing multi-species pattern.

son & Greig-Smith (1986) on the ground that the position of zero point in the ordination lacks ecological significance.

Traditionally, applications of the eigenanalysis-based methods discussed above have been almost completely confined to *transect* data: blocks are formed by successive fusions of neighbouring cells in the transect. Block shape is therefore not constant: duplication in size means duplication of the length/width ratio of the block. Consequently, one particular direction of variation will dominate the analysis: *characteristic distances* rather than *areas*, two different aspects of pattern (compositional vs directional), are revealed. Transect techniques will not be directly comparable to pattern analyses with information theory functions and expected resemblance which do not emphasize any particular direction. However, the ordination methods can just as well be applied to grids, a possibility which needs to be explored.

## V. Discussion

One objective of the present paper is to call attention of ecologists to scale problems in vegetation studies. We are not the first, of course, see e.g., Wiens (1989), for a more general presentation of the topic. Although it is simpler to rely on data obtained at a single scale point, we repeatedly emphasize that, no matter how sophisticated the subsequent analysis is, the success of any study of pattern-related vegetation phenomena is conditioned upon whether scale dependence is considered or not. We have to admit, at the same time, that realizing, accepting and adapting this requirement is difficult, and may prove too expensive and time-consuming. In vegetation dynamics, spatio-temporal patterns are analyzed, so that considering both temporal and spatial scaling simultaneously is a prerequisite, but relevant examples are few (Juhász-Nagy & Podani 1983, Czárán & Bartha 1989, Dale & Blundon 1990, Bartha 1993)

We point out that the information theory models described by Juhász-Nagy provide a sophisticated means for analyzing small-scale community behaviour and for the evaluation of the role of each constituent population in community structure. In fact, complex biological patterns can be described using community-, coalition- and single species-level approaches simultaneously, and are treated in a comprehensive way. The additivity of information theory estimates as applied to contingency tables offers an opportunity to decompose com-

munity-level estimates to lower level components without loss of any information. The methods described in this paper, however, illustrated only the fundamental concepts, there are many more derived functions for assessing, for example, *dis-sociatum* (i.e., the amount of information outside interaction) and so on. The more thorough discussion of this topic, however, was beyond the scope of this review (interested readers should consult Juhász-Nagy 1976, 1984). Although some of the methods (associatum and florula diversity estimation) require relatively large sample sizes, they are well worth using because of the clear theoretical foundation and fine details of results on various aspects of texture and interaction. The other procedures also have their relative merits, e.g., simplicity, ease of calculation, etc. They are recommended in (pilot) studies to detect approximate plot sizes that can be used in subsequent classifications, ordinations, or in studies of any functional community characteristics.

Juhász-Nagy's approach by itself suggests that there is no single scale point in plant communities which could be useful in general. Every function responds to different aspects (such as diversity or similarity) of community texture and structure, so it would be very naive to expect the existence of any single characteristic area. The relative positions of characteristic areas pertaining to different functions, the so-called characteristic ordering (Juhász-Nagy 1967), provide a more detailed exploration of communities. It also demonstrates that plant communities are more complex entities than assumed by any simplistic approach.

This review of methods may not be complete, of course, and we feel sorry if some important work was forgotten (or remained unknown for us) when preparing this paper. New methods may also appear in the vast literature of community ecology any time. We think that more emphasis should be placed in the future on comparisons between actual values and the corresponding random (null) reference. Detailed investigations are required to find analytical solutions in cases when it is possible at all. Monte Carlo simulations or the bootstrapping technique could be used otherwise. Comparisons between the diverging approaches discussed in this paper are also badly needed, since available comparative studies are restricted just to a few methods.

**Acknowledgements.** We are grateful to P. Juhász-Nagy for guidance, never-ceasing inspiration and for being with us. The project was supported by Hungarian National

Research Fund grant No. T2047 to J. P., and T. C.; No. T5316 and 936 to S. B., and Austrian Scientific Foundation grant No. P9072-BIO to J. P. and T. C.

## References

- Barkman, J. J. 1979. The investigation of vegetation texture and structure. In: M. J. A. Werger (ed.), *The study of vegetation*. pp. 125-160. Junk, The Hague.
- Barkman, J. J. 1989. A critical evaluation of minimum area concepts. *Vegetatio* 85:89-104.
- Bartha, S. 1991. Diversity processes during revegetation on dumps from strip coal-mining. *Symp. on Biological Diversity, Madrid*, pp. 180-181.
- Bartha, S. 1992. Preliminary scaling for multi-species coalitions in primary succession. *Abstracta Botanica* 16:31-41.
- Bartha, S. 1993. Spatial pattern development in primary succession on dumps from strip coal-mining in Hungary. PhD Thesis, Vácátót.
- Bouxin, G. & N. Gautier. 1982. Pattern analysis in Belgian limestone grasslands. *Vegetatio* 49:65-83.
- Camiz, S. 1993. STATIS ordinations vs the Juhász-Nagy models: The predictability of an exploratory tool. *Abstracta Botanica* 17:29-36.
- Camiz, S. & A. Gergely. 1990. An exploratory method for determining optimal plot size in plant community stands. *Abstracta Botanica* 14:83-108.
- Czárán, T. & Bartha, S. 1989. The effect of spatial pattern on community dynamics: a comparison of simulated and field data. *Vegetatio* 83:229-239.
- Dale, M. B. 1993. Bourne and boundary: walking a fine line. *Abstracta Botanica* 17:11-28.
- Dale, M. R. T. & Blundon, D. J. 1990. Quadrat variance analysis and pattern development during primary succession. *J. Vegetation Sci.* 1:153-164.
- Dietworst, P., E. van der Maarel & H. van der Putten. 1982. A new approach to the minimal area of a plant community. *Vegetatio* 50:77-91.
- Escoufier, Y. 1973. Le traitement des variables vectorielles. *Biometrics* 29:751-760.
- Galiano, E. F. 1983. Detection of multi-species patterns in plant populations. *Vegetatio* 53:129-138.
- Gibson, D. J. & P. Greig-Smith. 1986. Community pattern analysis: A method for quantifying community mosaic structure. *Vegetatio* 66:41-47.
- Gounot, M. & M. Calléja. 1962. Coefficient de communauté, homogénéité et aire minimale. *Bull. Serv. Carte Phytogeogr.* 7B:181-210.
- Greig-Smith, P. 1985. *Quantitative plant ecology*. 3rd ed. Blackwell, Oxford.
- Hill, M. O. 1973. The intensity of spatial pattern in plant communities. *J. Ecol.* 61:225-235.
- Juhász-Nagy, P. 1967. On some 'characteristic area' of plant community stands. *Proc. Colloq. Inf. theory, Bolyai Math. Soc., Debrecen*, pp. 269-282.
- Juhász-Nagy, P. 1976. Spatial dependence of plant populations. Part 1. Equivalence analysis (An outline of a new model). *Acta Bot. Hung.* 22:61-78.
- Juhász-Nagy, P. 1984. Spatial dependence of plant populations. Part 2. A family of new models. *Acta Bot. Hung.* 30:363-402.
- Juhász-Nagy, P. 1993. Notes on compositional diversity. *Hydrobiologia* 249:173-182.
- Juhász-Nagy, P. & J. Podani. 1983. Information theory methods for the study of spatial processes and succession. *Vegetatio* 51:129-140.
- Kenkel, N. C., P. Juhász-Nagy & J. Podani. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83:195-207.
- Lepš, J. 1990a. Can underlying mechanisms be deduced from observed patterns? In: Krahulec, F., Agnew, A. D. Q., Agnew, S. & Willems, J. H. (eds.), *Spatial processes in plant communities*. pp. 1-11. Academia, Praha.
- Lepš, J. 1990b. Comparison of transect methods for the analysis of spatial pattern. In: Krahulec, F., Agnew, A. D. Q., Agnew, S. & Willems, J. H. (eds.), *Spatial processes in plant communities*. pp. 71-82. Academia, Praha.
- McCulloch, C. E. 1985. Variance tests for species association. *Ecology* 66:1676-1681.
- Moravec, J. 1973. The determination of the minimal area of phytocoenoses. *Folia Geobot. Phytotax.* 8:23-47.
- Noy-Meir, I. & D. J. Anderson. 1971. Multiple pattern analysis, or multiscale ordination: towards a vegetation hologram? In: G. P. Patil, E. C. Pielou & W. E. Waters (eds.), *Many species populations, ecosystems and systems analysis*. Vol. 3. *Statistical ecology*, pp. 207-231. Penn. State Univ. Press, University Park, PA, USA.
- Orlóci, L. 1988. Detecting vegetation patterns. *ISI Atlas of Science, Plants & Animals* 1:173-177.
- Økland, R. H., O. Eilertsen & T. Økland. 1990. On the relationship between sample plot size and beta diversity in boreal coniferous forests. *Vegetatio* 87:187-192.
- Palmer, M. W. 1987. Variability in species richness within Minnesota oldfields: a use of the variance test. *Vegetatio* 70:61-64.
- Peet, R. K. 1974. The measurement of species diversity. *Ann. Rev. Ecol. Syst.* 5:285-307.
- Pielou, E. C. 1966. Species-diversity and pattern-diversity in the study of ecological succession. *J. theor. Biol.* 10:370-383.
- Pielou, E. C. 1972.  $2^k$  contingency tables in ecology. *J. theor. Biol.* 34:337-352.
- Pielou, E. C. 1975. *Ecological diversity*. Wiley, New York.
- Pielou, E. C. 1977. *Mathematical ecology*. Wiley, New York.
- Podani, J. 1984. Analysis of mapped and simulated vegetation patterns by means of computerized sampling techniques. *Acta Bot. Hung.* 30:403-425.
- Podani, J. 1987. Computerized sampling in vegetation studies. *Coenoses* 2:9-18.

- Podani, J. 1992. Space series analysis of vegetation: processes reconsidered. *Abstracta Botanica* 16:25-29.
- Schluter, D. 1984. A variance test for detecting species associations, with some example applications. *Ecology* 65:998-1005.
- Sterling, A., B. Peco, M. Casado, E. F. Galiano & F. D. Pineda. 1984. Influence of microtopography on floristic variation in the ecological succession in grassland. *Oikos* 42:334-342.
- Tóthmérész, B. & Zs. Erdei. 1992. The effect of species dominance on information theory characteristics of plant communities. *Abstracta Botanica* 16:43-47.
- van Hoef, J. M. & D. C. Glenn-Levin. 1989. Multiscale ordination: a method for detecting pattern at several scales. *Vegetatio* 82:59-67.
- Watkins, A. J. & Wilson, J. B. 1992. Fine-scale structure of lawns. *J. Ecol.* 80:15-24.
- Whittaker, R. H. (ed.) 1973. *Ordination and classification of communities*. Junk, The Hague.
- Wiens, J. A. 1989. Spatial scaling in ecology. *Funct. Ecol.* 3:385-397.
- Wildi, O. & B. O. Krüsi. 1992. Long-term monitoring: the function of plot size and sampling design. *Abstracta Botanica* 16:7-14.
- Williams, W. T., G. N. Lance, L. J. Webb, J. G. Tracey & J. H. Connell. 1969. Studies in the numerical analysis of complex rain-forest communities. IV. A method for the elucidation of small-scale forest pattern. *J. Ecol.* 57:635-654.
- Wilson, J. B., Gitay, H. & Agnew, A. D. Q. 1987. Does niche limitation exist? *Funct. Ecol.* 1:391-397.
- Wilson, M. V. & A. Shmida. 1984. Measuring beta diversity with presence-absence data. *J. Ecol.* 72:1055-1064.
- Zobel, K., Zobel, M. & Peet, R. K. 1993. Change in pattern diversity during secondary succession in Estonian forests. *J. Vegetation Sci.* 4:489-498.